https://doi.org/10.4038/sljma.v27i1.5610

The Sri Lanka Journal of Medical Administration

https://sljma.sljol.info/ Original Article Volume 27 Issue 01

Evaluation of Ambient AI-Assisted Clinical Documentation and Decision Support: An 11-Week Service Evaluation at a UK Teaching Hospital

Rajapaksha D M P[1]*, Jeewarathne K H P[2]

[1]Base Hospital, Deniyaya, [2]Base Hospital, Elpitiya

Correspondence to:
Rajapaksha D M P

E-Mail
malalarajapaksha7@gmail.com

Key Words:

Ambient AI; Artificial Intelligence; Clinical Documentation; Electronic Patient Record; Clinical Decision Support; Digital Health

Abstract

Background: Ambient artificial intelligence (AI)-assisted listening systems may reduce clinical documentation burden and support more efficient clinical workflows. Liverpool University Hospitals NHS Foundation Trust (LUHFT) piloted ambient AI technology across selected clinical services.

Objective: To evaluate the operational and documentation-related effects of ambient AI-assisted listening and explore the inclusion of final clinical diagnoses within AI-generated initial differential diagnosis lists.

Methods: An 11-week service evaluation was conducted across colorectal outpatient clinics, Accident and Emergency (A&E) triage, cardiology clinics, and diabetes clinics. A total of 402 de-identified patient encounters were evaluated: colorectal (n=98), A&E (n=99), cardiology (n=104), and diabetes (n=101). Data were obtained through the LUHFT Business Intelligence (BI) system under institutional governance arrangements. Documentation time was derived from electronic timestamps. Final-diagnosis inclusion within the initial AI-generated differential diagnosis was assessed descriptively. Clinician and digital programme discussions provided qualitative feedback on documentation and workflow.

Results: Documentation time ranged from 12–20 minutes before implementation and 8–15 minutes during ambient AI-assisted documentation. The final clinical diagnosis was represented within the initial AI-generated differential diagnosis list in 98% of evaluated encounters. This represents diagnostic-list inclusion rather than diagnostic accuracy. Clinicians perceived AI-assisted documentation as more structured and reported improved focus on patient interaction. Overall waiting times were lower during the implementation period; however, concurrent patient-flow interventions prevented attribution of this change specifically to ambient AI.

Conclusion: Ambient AI-assisted listening was associated with a lower observed documentation-time range and favourable clinician perceptions of documentation and workflow. Further prospective evaluation of diagnostic performance, safety, patient experience, and cost-effectiveness is required.

Introduction

Healthcare systems face increasing pressure from rising patient demand, workforce shortages, and expanding administrative and clinical documentation requirements. [1,2] Although Electronic Patient Record (EPR) systems improve accessibility and continuity of clinical information, extensive data entry may increase clinician workload and reduce time available for direct patient interaction.

Advances in artificial intelligence (AI), natural language processing, and speech recognition have enabled ambient listening systems that capture clinician–patient conversations and generate structured draft clinical documentation. [3,4,5] Some systems also provide clinical decision-support functions, including differential diagnoses, suggested investigations, and guideline-informed management options. [6]

Liverpool University Hospitals NHS Foundation Trust (LUHFT) introduced an ambient listening pilot as part of its digital transformation programme. The system generated structured clinical documentation and supported differential diagnosis and clinical decision-making, while clinicians retained responsibility for reviewing AI-generated outputs and making final clinical decisions.

This service evaluation examined its effects on documentation and workflow and explored the

extent to which final clinical diagnoses were represented within initial AI-generated differential diagnosis lists.

Justification

Clinical documentation is essential for patient safety, continuity of care, communication, and medico-legal accountability, but it contributes substantially to clinician workload. Ambient AI may reduce repetitive documentation while allowing greater attention to patients.

Real-world evaluation is required because AI implementation also raises concerns regarding reliability, patient safety, privacy, information governance, and over-reliance on automated recommendations. [7]

The experience is relevant to medical administration and potentially to Sri Lanka, where high patient volumes, workforce limitations, and increasing digitalisation create a need for efficient clinical information systems.

Objectives

General Objective

To evaluate the operational and documentation-related effects of ambient AI-assisted listening within clinical workflows and explore its potential contribution to clinical decision support.

Specific Objectives

- To assess changes in documentation time associated with ambient AI-assisted listening.
- To explore clinician perceptions of documentation quality, patient interaction, and workflow.
- To determine the proportion of evaluated encounters in which the final clinical diagnosis was represented within the initial AI-generated differential diagnosis list.
- To describe the role of ambient AI within the EPR-enabled clinical pathway.
- To consider implications for future digital health development in Sri Lanka.

Literature Review

Ambient clinical AI combines speech recognition, natural language processing, and generative AI to convert clinical conversations into structured documentation. Emerging evidence suggests that digital scribes and ambient AI may reduce documentation burden and improve clinician experience. [3,4,5]

Reducing computer interaction during consultations may enable clinicians to maintain greater engagement with patients. [8] More advanced systems may also support clinical reasoning by suggesting differential diagnoses, investigations, and management options. [6]

However, AI-assisted decision support must be distinguished from autonomous diagnosis. The inclusion of a correct diagnosis within a differential list does not establish diagnostic accuracy, particularly when the list contains multiple possibilities. AI-generated outputs may also be inaccurate or incomplete. Human oversight, information governance, confidentiality, and accountability therefore remain essential. [7]

Evidence concerning ambient clinical AI remains limited, particularly in South Asian healthcare systems. [9] Real-world service evaluations can therefore provide useful information regarding implementation and future research priorities.

Methodology

Study Design and Setting

An 11-week service evaluation of ambient AI-assisted listening was conducted at LUHFT across four clinical settings: colorectal outpatient clinics, A&E triage, cardiology clinics, and diabetes clinics.

Approximately 100 encounters from each clinical area were requested through the LUHFT BI system. The system-generated samples varied slightly, producing 402 de-identified patient encounters: colorectal (n=98), A&E (n=99), cardiology (n=104), and diabetes (n=101).

The patient encounter was the unit of analysis. Individual clinician numbers were not extracted because patients could be managed by multiple clinicians during their care pathway.

Ambient AI Workflow

During consultations, the ambient listening system captured clinician–patient conversations and generated structured clinical documentation. It also provided decision-support functions, including differential diagnoses and suggestions regarding investigations and management.

Clinicians performed physical examinations, entered relevant findings, interpreted investigation results, and reviewed AI-generated outputs. Final diagnostic and management decisions remained the responsibility of treating clinicians.

Data Collection and Outcome Measures

Operational data were obtained through the LUHFT BI system under institutional governance arrangements. Data were provided in de-identified form, and identifiable patient information was not exported from the secure LUHFT environment.

Documentation time was derived from routinely captured electronic timestamps representing the interval between commencement and submission/completion of documentation. Available data were reported as ranges and were therefore analysed descriptively.

Final-diagnosis inclusion was defined as the presence of the final clinical diagnosis anywhere within the initial AI-generated differential diagnosis list. The proportion was calculated against all encounters evaluated for this outcome. This was an exploratory measure of diagnostic-list inclusion and not a measure of diagnostic accuracy.

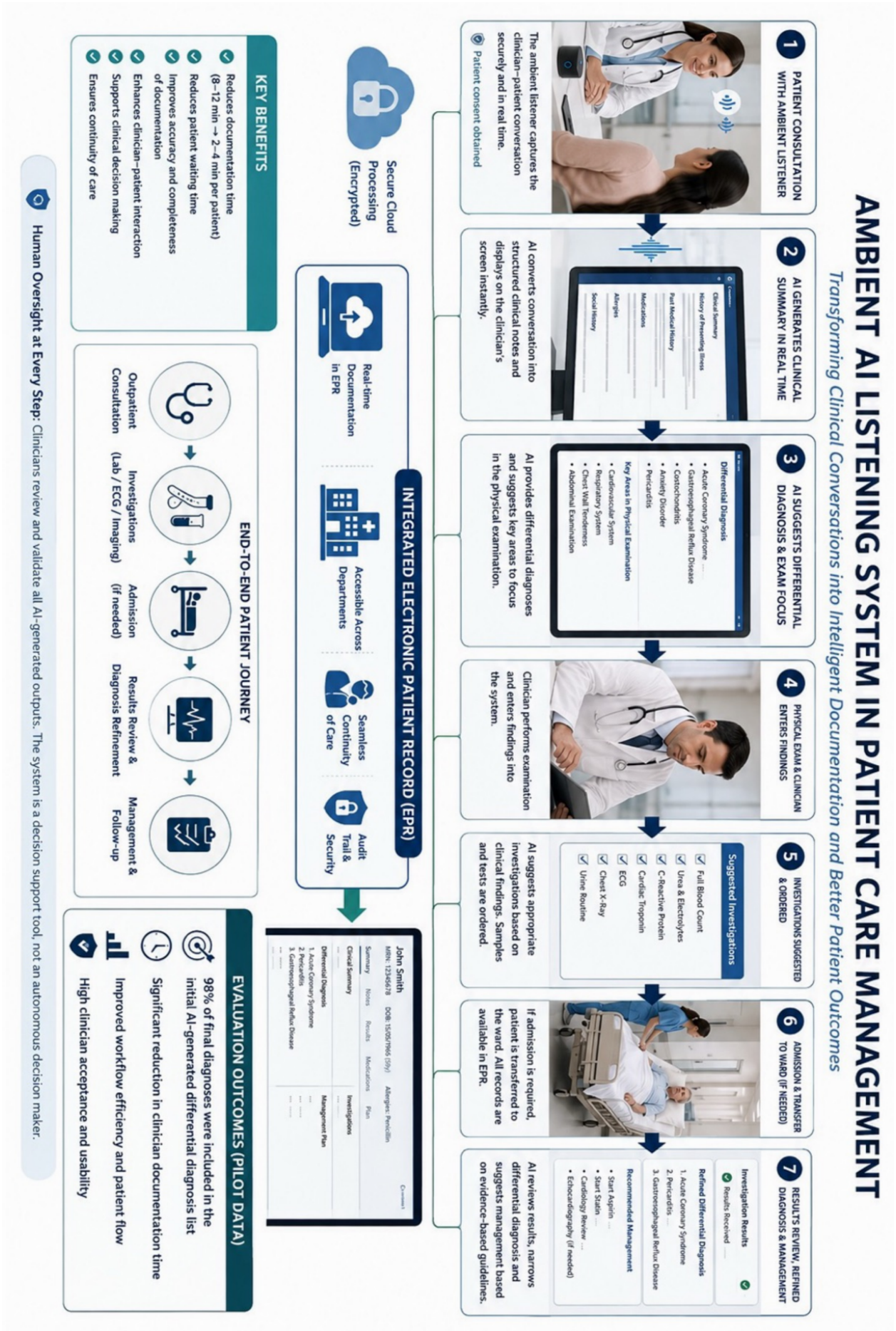


Figure 1: Ambient AI-assisted clinical documentation and decision-support workflow integrated with the Electronic Patient Record system, with clinician oversight throughout

Perceptions of documentation quality, workflow, and clinician–patient interaction was obtained through clinician discussions and digital programme group meetings. No validated quantitative documentation-quality instrument was used; these findings were therefore treated as qualitative stakeholder perceptions.

Patient waiting time was routinely monitored by the hospital. However, because other patient-flow interventions occurred concurrently, changes in waiting time were not attributed specifically to ambient AI.

Incorrect or potentially unsafe AI-generated recommendations were not systematically quantified. AI-generated outputs remained subject to clinician review before clinical use.

Ethical and Governance Considerations

The project was undertaken as a service evaluation within LUHFT institutional governance arrangements. Applicable patient consent, confidentiality, and information-governance requirements were followed. Evaluation data were provided through authorised institutional processes in de-identified form.

Results

A total of 402 patient encounters were evaluated: colorectal outpatient clinics (98; 24.4%), A&E triage (99; 24.6%), cardiology clinics (104; 25.9%), and diabetes clinics (101; 25.1%).

Documentation time 12–20 min before; 8–15 min during ambient AI Lower observed documentation-time range

Final diagnosis included in initial AI differential 98% of evaluated encounters Diagnostic-list inclusion; not diagnostic accuracy

Documentation quality Perceived as more structured and of better quality Qualitative stakeholder finding

Clinician–patient interaction Greater focus on direct patient interaction perceived Qualitative stakeholder finding

Waiting time Lower during implementation period Cannot be attributed specifically to ambient AI

The documentation-time range decreased from 12–20 minutes before implementation to 8–15 minutes during ambient AI-assisted documentation. No inferential statistical testing was undertaken.

The final clinical diagnosis was represented within the initial AI-generated differential diagnosis list in 98% of evaluated encounters. The rank of the diagnosis within the list, including top-one, top-three, or top-five performance, was not evaluated.

Clinician and digital programme discussions indicated that AI-assisted documentation was perceived as more structured and of better quality and enabled greater focus on patient interaction. These were qualitative perceptions rather than objectively measured improvements.

Waiting times were lower during the implementation period; however, several concurrent patient-flow initiatives prevented causal attribution to ambient AI.

Discussion

This service evaluation provides real-world evidence regarding ambient AI-assisted clinical documentation across four clinical settings in a large UK teaching hospital.

The principal quantitative operational finding was a

Table 01: Principal Findings

Outcome	Finding	Interpretation
Documentation time	12–20 min before; 8–15 min during ambient AI	Lower observed documentation-time range
Final diagnosis included in initial AI differential	98% of evaluated encounters	Diagnostic-list inclusion; not diagnostic accuracy
Documentation quality	Perceived as more structured and of better quality	Qualitative stakeholder finding
Clinician–patient interaction	Greater focus on direct patient interaction perceived	Qualitative stakeholder finding
Waiting time	Lower during implementation period	Cannot be attributed specifically to ambient AI

reduction in the observed documentation-time range from 12–20 minutes to 8–15 minutes. Although the available range data did not permit inferential statistical testing, the finding suggests potential for reducing documentation burden.

Clinicians also perceived documentation as more structured and reported greater ability to focus on patient interaction. Similar potential benefits have been reported with ambient AI and digital-scribe technologies. [3,4,5] However, these findings should be interpreted as stakeholder perceptions because no validated documentation-quality or clinician-experience instrument was used.

The final clinical diagnosis appeared within the initial AI-generated differential diagnosis list in 98% of evaluated encounters. This finding is potentially encouraging but must be interpreted carefully. It demonstrates final-diagnosis inclusion, not 98% diagnostic accuracy. The rank and number of diagnoses within the differential lists were not evaluated. A broad differential list could achieve high inclusion without equivalent diagnostic specificity.

Future studies should therefore assess top-one, top-three, and top-five performance and systematically examine incorrect or potentially unsafe recommendations. Measures such as sensitivity, specificity, and predictive values would be required before conclusions regarding diagnostic performance could be made.

Clinical oversight remained central to the workflow. The technology supported rather than replaced professional judgement, with clinicians retaining responsibility for examination, interpretation of investigations, validation of AI-generated information, diagnosis, and management.

Lower waiting times were observed during implementation, but multiple patient-flow interventions were occurring simultaneously. Consequently, the evaluation cannot establish that ambient AI caused the observed reduction.

Implications for Sri Lanka

Ambient AI may have relevance to Sri Lankan hospitals, particularly high-volume outpatient departments, emergency treatment units, and specialist clinics. Potential benefits include reduced documentation burden, more structured clinical records, and improved continuity of information.

However, implementation would require adequate digital infrastructure, interoperability, information-security safeguards, and clear governance arrangements. Sinhala, Tamil, English, and mixed-language clinical conversations would require local validation.

A phased pilot approach in selected digitally enabled hospitals would therefore be preferable before wider adoption, with evaluation of documentation accuracy, safety, clinician and patient experience, operational efficiency, and cost-effectiveness.

Limitations

This was an 11-week service evaluation within selected clinical areas of a single NHS Trust. Documentation-time data were available as ranges, limiting statistical analysis. Documentation quality and clinician experience were based on stakeholder discussions rather than validated instruments.

The 98% finding represents final-diagnosis inclusion within the AI-generated differential list rather than diagnostic accuracy, and diagnosis ranking was not evaluated. Incorrect or potentially unsafe AI recommendations were not systematically quantified.

Waiting-time changes could not be attributed independently to ambient AI because other interventions occurred concurrently. Long-term patient outcomes and cost-effectiveness were not evaluated.

Conclusion

This 11-week service evaluation involving 402 patient encounters demonstrated a lower observed documentation-time range during ambient AI-assisted documentation. The final clinical diagnosis appeared within the initial AI-generated differential diagnosis list in 98% of evaluated encounters, representing a high final-diagnosis inclusion rate rather than diagnostic accuracy.

Clinicians reported favourable perceptions regarding documentation and patient interaction. Ambient AI therefore shows potential to reduce documentation burden and support digitally integrated clinical workflows, but larger prospective evaluations incorporating validated outcome measures, diagnostic-performance assessment, safety monitoring, patient experience, and economic evaluation are required.

The experience provides potentially useful lessons for future evaluation of ambient AI within Sri Lanka's evolving digital health system.

Recommendations

Future evaluations should use patient-level documentation-time data and validated measures of documentation quality, clinician experience, and patient experience. Diagnostic-support studies should assess top-one, top-three, and top-five performance and systematically record incorrect or potentially unsafe AI outputs. Formal economic evaluation should also be undertaken.

For Sri Lanka, carefully governed pilot implementation should initially be considered in digitally enabled hospitals, with particular attention to interoperability, data governance, clinician oversight, and performance across Sinhala, Tamil, and English.

References

1. Danaher. Improving how clinicians communicate with patients: an integrative review and framework. Journal of Service Research. 2023.
2. Glarcher. Addressing nurse shortages and pandemic responses to enhance patient safety within global health challenges. Journal of Advanced Nursing. 2025.
3. Kanaparthy. Real-world evidence synthesis of digital scribes using ambient listening and generative artificial intelligence for clinician documentation workflows: rapid review. JMIR. 2025.
4. Owens. Effect of ambient voice technology, natural language processing, and artificial intelligence on the patient–physician relationship. Applied Clinical Informatics. 2024.
5. Wang. Evaluating the performance of artificial intelligence-based speech recognition for clinical documentation: a systematic review. BMC Medical Informatics and Decision Making. 2025.
6. Khurshid. The Impact of AI-Powered Clinical Decision Support Systems on Clinical Decision-Making and Treatment Quality: A Systematic Review. Journal of Islamabad Medical & Dental College. 2025.
7. World Health Organization (WHO). Ethics and governance of artificial intelligence for health: large multi-modal models. WHO guidance. World Health Organization; 2024.
8. Roche. Patient and clinician perceptions of factors relevant to ideal specialty consultations. JAMA Network. 2022.
9. Aldosari. The use of digital health by South Asian communities: scoping review. Journal of Medical Internet Research. 2025.